



# INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

## KNOWLEDGE DISCOVERY PROCESS IN DATA MINING: A REVIEW

SUNIL SAINI

Assistant Professor, Deptt. of Computer Sc. & Engg., Kurukshetra Institute of Technology & Management, Kurukshetra, Haryana.

Accepted Date: 15/05/2017; Published Date: 01/06/2017

**Abstract:** - This era is data driven; Internet is ubiquitous in this era. Due to which each individual is generating data from different sources. From statistics about 2.5 quintillion bytes of data is created on daily basis. From many years different companies have been developing data warehouse to store data and for analyzing the data from different perspectives. This analysis brings confidence in analyst to make wise decisions for their organization. Data analytics improve performance and profit of an organization. For knowledge discovery data from different sources are captured like social networks, sensors data, online news, climate data, archived data and so on. Due to richness of new data sources there is a need for knowledge discovery. This paper provides a review of knowledge discovery process in data mining.

**Keywords:** Cleaning, Data-mining, Evaluation, Integration



PAPER-QR CODE

Corresponding Author: SUNIL SAINI

Access Online On:

[www.ijpret.com](http://www.ijpret.com)

How to Cite This Article:

Sunil Saini, IJPRET, 2017; Volume 5 (10): 15-19

## INTRODUCTION

Knowledge discovery is defined as a process of extracting useful, novel and valid knowledge from raw data. Due to data driven era, loads of data is accumulating, which needs analysis. This analytic process is useful in finding relationships between data items, finding hidden patterns and searching useful insights which will be useful in optimizing the process of an organization and for earning more profits. On daily basis 2.5 quintillion bytes of data is created [1]. Data mining is useful in various areas like databases, statistics, machine learning, optimization, data visualization and many more. The abundance of data makes knowledge discovery process, an important part of research and attracts many researchers. Data mining is also termed as knowledge mining, pattern analysis, knowledge extraction and many more.

## II. STEPS IN Knowledge Discovery process

Data is fuel to knowledge discovery process. If data is not available then mining is not possible. Data has to be preprocessed before it will be used in mining process. Cleaning, integration of data, selection of data, transformation of data are different steps which are followed in data preprocessing and after these steps data is ready for mining [2] [7].

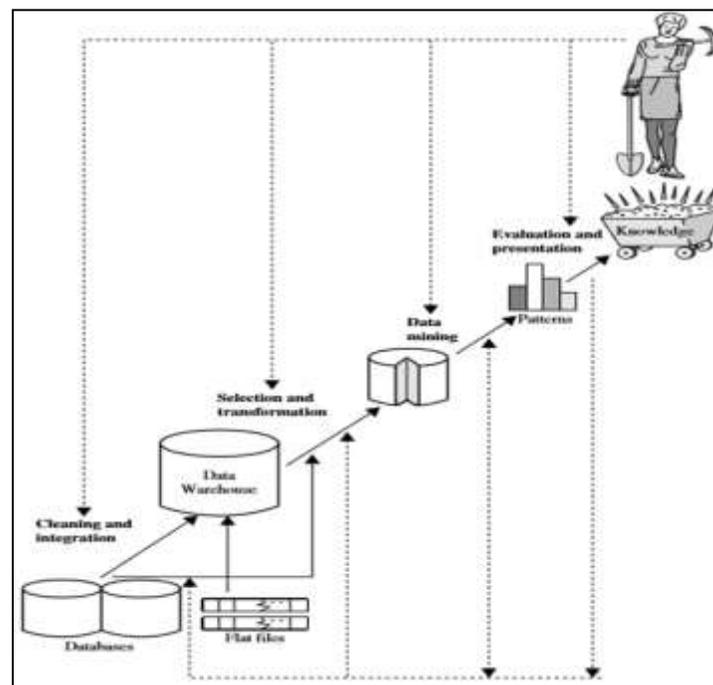


Figure: Knowledge Discovery Process

Different steps of knowledge discovery process is shown in figure [7]

### **A. Cleaning**

Cleaning is the process of removal of irrelevant data, inconsistent data. Example of inconsistent data is errors while data entry, redundancies etc. In this process missing values are identified, either those values are removed or values are written on the basis of attribute mean or fill missing value manually etc. Noisy data is defined as a data having meaningless information in it known as noise. It is also known as corrupted data. Noisy data is handled with the help of binning methods, clustering, regression etc [2][3]

### **B. Integration of Data**

For mining data is collected from heterogeneous sources and then combined in a common place from where mining process can be applied. Different source can be data warehouses, databases, flat files etc. [2] [3].

### **C. Selection of Data**

Selection is a process of retrieving relevant data from database, data warehouse. Relevance of data is decided in accordance to problem definition and only that data is selected which is useful for problem [2] [6].

### **D. Transformation of Data**

Another name for transformation is consolidation. In this process data is transformed into a form which is appropriate for data mining. Different data transformation techniques are normalization, smoothing, aggregation, generalization etc. [2] [3].

### **E. Data Mining**

With the help of data mining process, actionable information or patterns are discovered from huge data sets. Mathematical analysis is used to derive patterns and insights from the data. Data mining can be used various scenarios like estimating sales, for finding sequences, for recommendations to customers or sellers and in many more scenarios. It can be applied in various disciplines like statistics, machine learning, visualization, information science etc. [2] [4]

### **F. Evaluation & Knowledge Presentation**

Evaluation means understanding of results, check whether the discover patterns are interesting and novel, impact of the discovered information and patterns, result interpretation by domain experts. Interesting patterns are those which are easily understood by humans and they are

also valid and useful. Mined knowledge or patterns have to be presented to the user with the help of bar charts, pie-charts, histograms, graphs for better interpretation of results [2] [5].

### III. TOOLS FOR DATA MINING

There are various tools for data mining which makes the mining process very easy. Some of them are listed below [9]

- WEKA- Machine Learning & Data Mining
- RapidMiner
- KNIME
- Orange
- R Language
- IBM Intelligent Miner
- Statistica- Statsoft
- Data Miner and many more

### IV. CONCLUSION

Data mining is very hot topic in this era. Data mining is used to find useful information and patterns which are useful for an organization for improving their working style. With the help of this paper, a brief introduction to knowledge discovery process is given with brief introduction to its different steps. List of different data mining tools is also presented. With this paper researchers will be able to know knowledge discovery process easily.

### REFERENCES

1. <https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/>
2. Jiawei Han and Micheline Kamber : "Data Mining: Concepts and Techniques". 3<sup>rd</sup> Edition.
3. Malik, Jasdeep Singh, Prachi Goyal, and Akhilesh K. Sharma. "A Comprehensive Approach towards Data Preprocessing Techniques & Association Rules." In Proceedings of The4th National Conference. 2010.

4. Priyadharsini.C, Dr. Antony Selvadoss Thanaman," An Overview of Knowledge Discovery Database and Data mining Techniques", International Journal of Innovative Research in Computer and Communication Engineering, 2014
5. Database and Data mining Techniques", International Journal of Innovative Research in Computer and Communication Engineering, 2014
6. Data mining knowledge representation,"[http://www.cs.ccsu.edu/~markov/ccsu\\_courses/datamining-4.pdf](http://www.cs.ccsu.edu/~markov/ccsu_courses/datamining-4.pdf)"
7. Data mining: Concepts and Techniques,"  
[http://www.vssut.ac.in/lecture\\_notes/lecture1422914558.pdf](http://www.vssut.ac.in/lecture_notes/lecture1422914558.pdf)"
8. Cios K.J., Pedrycz, W. ,Swiniarski, R.W.,Kurgan L," The Knowledge Discovery Process", Springer, 2007
9. Knowledge Discovery Process and Data Mining  
<http://www.cs.put.poznan.pl/jstefanowski/sed/DM12-summary.pdf>